

VOICE CONTROLLER AND VOICE-CONTROLLER SYSTEM HAVING A VOICE-  
CONTROLLED APPARATUS

Background of the Invention:

5 Field of the Invention:

3767800-01001  
The present invention relates to a voice controller utilizing a sound detector detecting a sound signal containing a voice command and to a system including such a voice-controlled apparatus.

10 Human speech is an expedient way to control television sets or other items of electronic entertainment equipment. Using voice-controlled man-machine interfaces provide many advantages including simpler operation of the respective item of equipment for the user. Therefore, retrofitting a voice  
15 control system is particularly beneficial to existing equipment that needs a voice control option.

The problem with using voice control of audio equipment is that the sound or audio signal produced by the equipment (typically a loudspeaker) is mixed with the spoken voice  
20 command and thus superimposed over the latter. This overlap worsens the recognizability of the voice command for the voice

recognition system. The same also applies to external sound sources, whose noise or audio signals can be superimposed on the voice command of the user. If, for example, a television set is to be voice controlled, and if a stereo system or the like in the same room is switched on, the voice command (the signal) is considerably less recognizable because the audio signals produced by the stereo system (the noise) are superimposed over the spoken command of the user. The unknown interfering noises added to the voice command creates a resulting sound signal that exhibits a relatively poor signal-to-noise ratio in the resulting sound signal. The resulting sound signal is supplied to the voice recognition system of the voice-controlled item of equipment where the resulting sound signal is to be converted into a corresponding control signal. The poor signal-to-noise ratio inhibits the conversion.

#### Summary of the Invention:

It is accordingly an object of the invention to provide a voice-controller and a voice-controller system having a voice-controlled apparatus that overcomes the hereinafore-mentioned disadvantages of the heretofore-known devices of this general type and that ensures the recognizability of a voice command with a sufficiently good signal-to-noise ratio, even in the

presence of a sound source whose sound signal is superimposed over the voice command.

With the foregoing and other objects in view, there is provided, in accordance with the invention, a voice controller including a sound source, a sound detector, a receiver, and a sound signal processor. The sound source includes a transmitter. The sound detector detects a sound signal containing a voice command. The sound detector has a voice recognizer recognizing the voice command. The sound detector converts the voice command into a corresponding control signal for a voice-controlled apparatus. A receiver receives sound information from the transmitter associated with the sound source. A sound signal processor coupled to the sound detector and the receiver. The sound signal processor corrects the sound signal by eliminating the sound information from the sound signal to produce a corrected sound signal, and supplies the corrected sound signal to the voice recognizer for evaluation.

In accordance with another feature of the invention, the sound detector, the receiver, the sound signal processor, and the voice recognizer are arranged in a mobile part provided separately from said voice-controlled apparatus.

In accordance with another feature of the invention, the voice-controlled apparatus includes a voice-controller receiver, and the mobile part has a transmitter transmitting the corresponding control signal to the voice-controller receiver.

In accordance with another feature of the invention, the transmitter of the mobile part communicates with the voice-controller receiver via a wireless communication channel.

In accordance with another feature of the invention, the sound signal processor determines a degree of correlation between the sound signal detected by the sound detector and a sound signal corresponding to the sound information. The sound signal processor determines an acoustic delay between the sound signal detected by the sound detector and a sound signal corresponding to the sound information. The sound signal processor corrects the sound signal detected by said sound detector while accounting for the acoustic delay.

In accordance with another feature of the invention, the sound signal processor determines the degree of correlation between the sound signal detected by cross-correlating the sound detector and the sound signal corresponding to the sound information.

In accordance with another feature of the invention, the sound signal processor subtracts the sound signal corresponding to the sound information from the sound signal detected by the sound detector, while accounting for the determined acoustic delay, to obtain a corrected sound signal to be supplied to the sound signal processor.

In accordance with another feature of the invention, the sound detector includes a number of microphones that are coupled to one another. The microphones have an acoustic phase shift between them. And, the sound detector accounts for the acoustic phase shift present between the number of microphones.

In accordance with another feature of the invention, the voice controller includes a keyboard in the sound detector. The keyboard programs the voice recognizer.

In accordance with another feature of the invention, the sound signal processor is associated with a number of sound sources, and the sound signal processor separately corrects for each of the number of sound sources.

With the objects of the invention in view, there is also provided a voice-controller system. The voice-controller system includes a voice controller as described above. In addition, the voice-controller system includes a receiver  
5 receiving sound information from a transmitter associated with a sound source, and a sound signal processor coupled to the sound detector and the receiver. The sound signal processor corrects the sound signal by eliminating the sound information from the sound signal to produce a corrected sound signal, and  
10 supplies the corrected sound signal to the voice recognizer for evaluation. The voice-controller system includes a sound source associated with a transmitter transmitting the sound information to the receiver of the voice-controlled apparatus. The sound information in each case describes the sound signal  
15 generated by the sound source.

In accordance with a further feature of the invention, the transmitter associated with the sound source communicates with the receiver associated with the voice-controlled apparatus via a wireless communication channel.

20 In accordance with a further feature of the invention, the wireless communication channel is an infrared channel. The wireless communication channel also can be a radio channel.

In accordance with a further feature of the invention, the voice-controlled apparatus itself belongs to the at least one sound source, so that the sound information transmitted by the transmitter to the receiver associated with the voice-  
5 controlled apparatus describes the sound signal generated by the voice-controlled apparatus at that instant.

In accordance with a further feature of the invention, the voice-controlled apparatus is an item of electronic entertainment equipment.

10 In accordance with a further feature of the invention, the at least one sound source is an item of electronic entertainment equipment.

According to the invention, the voice controller is assigned a receiver that receives sound information from transmitter that  
15 are associated with at least one sound source. This sound source can be, for example, a loudspeaker belonging to the apparatus itself or else loudspeakers belonging to other equipment. The voice controller according to the invention is accordingly preferably used in a system having at least one  
20 sound source. This sound source can be associated with transmitting means for transmitting the sound information to the receiving means of the voice controller. The sound

information in each case describes the sound signal generated by the sound source at that instant.

The sound information is used to communicate, in particular, the pitch, loudness etc. of the audio or sound signal produced by the corresponding sound source at that instant.

The voice controller receives a voice command in the form of a sound signal. However, this sound signal is composed not only of the voice command but also of the surrounding noise. Usually, the surrounding noise is the audio signal produced by the sound source. Because the audio signal respectively generated at that instant is known to the voice controller, on account of the sound information, the sound signal registered by the voice controller can be corrected appropriately and freed of that component that corresponds to the audio signal from the sound source. The voice recognition is based only on the sound signal corrected or filtered in this way.

Because the corrected sound signal contains only still unknown interfering noise in addition to the voice command, the corrected sound signal has a considerably improved signal-to-noise ratio (S/N) as compared with the original sound signal.



The transmission of the sound information to the receiver associated with the voice controller can, in particular, be carried out in a wireless manner, for example via an infrared or radio channel.

5 It is particularly advantageous if these receivers are integrated into a mobile part (remote control) provided to operate the voice controller. Likewise, the sound signal processors that carry out the above-described correction or filtering of the sound signal including the voice command, and also the voice recognizer that carry out the subsequent voice recognition, can be integrated into the mobile part. Then, by means of voice recognition, the voice recognizer generate a control signal that corresponds to the corrected sound signal and that, for example, is transmitted to the voice controller  
10 via an infrared transmitter. However, the receiver and the sound signal processor can be located in the voice controller itself.  
15

The present invention is generally suitable for the voice control of apparatus of any desired configuration, in particular for the voice control of electronics entertainment equipment, such as stereo systems or television sets.  
20

Other features that are considered as characteristic for the invention are set forth in the appended claims.

Although the invention is illustrated and described herein as embodied in a voice controller and voice-controller system  
5 having such a voice-controlled apparatus, it is nevertheless not intended to be limited to the details shown, since various modifications and structural changes may be made therein without departing from the spirit of the invention and within the scope and range of equivalents of the claims.

10 The construction and method of operation of the invention, however, together with additional objects and advantages thereof will be best understood from the following description of specific embodiments when read in connection with the accompanying drawings.

15 Brief Description of the Drawings:

Fig. 1 is a diagrammatic view showing a voice controller according to the invention in a system having a number of sound sources; and

Fig. 2 is a schematic view depicting the construction of the  
20 remote control shown in Fig. 1.

Description of the Preferred Embodiments:

In all the figures of the drawing, sub-features and integral parts that correspond to one another bear the same reference symbol in each case.

5 Referring now to the figures of the drawings in detail and first, particularly to Fig. 1 thereof, there is shown a stereo system having an amplifier 22 and a number of loudspeakers 24. Various audio or sound sources 20, such as a tuner, a cassette deck, a DAT device, or a video recorder, are connected to the  
10 amplifier 22 via connecting leads 23. In addition, in the example illustrated, a television set 21 is connected to the amplifier 22.

An additional device 16 is connected to the amplifier 22 via a connecting lead 19. The additional device 16 also can be  
15 integrated in one of the items of equipment illustrated. The audio signals generated by the entire system, that is to say the television set 21 and the stereo system 20, 23, are received by the additional device 16 via the connecting lead 19, are mixed together, encoded and/or modulated and converted  
20 into corresponding analogue or digital sound or audio information which, with the aid of a transmitter 18, is transmitted via a transmission channel 8 to a corresponding receiver 5 of a remote control 1. The transmission channel 8

can be, in particular, an infrared or radio channel, for example one meeting the Bluetooth mobile radio standard. On the other hand, control information for the operation of the television set 21 or the stereo system 20, 22, 23 is transmitted to a receiver 17 of the additional device 16 by a transmitter 6 in the remote control 1 via a transmission channel 7. The additional device 16 is therefore connected bidirectionally to the remote control 1.

Fig. 2 shows the remote control 1. The remote control 1 includes, for the case of infrared transmission, an IR receiving diode as receiver 5 with an IR receiver amplifier/IR converter 15 connected downstream that supplies the sound information received from the additional device to a processor 9.

In addition, the remote control 1 includes at least one microphone 4. The microphone 4 picks up the sound signal respectively acting on the remote control at that instant. The sound signal that is picked up by the microphone 4 contains a voice command from a user. An example of a voice command could be to activate the television set 21 or the stereo system 20, 22, 23. Then, the sound signal is supplied via one or more receiving amplifiers 10 to an analogue/digital

converter 11 and thus digitized. The digitized sound signal is finally supplied to the processor 9 for voice recognition.

The operation of the processor 9 is carried out on the basis of a program stored in a program memory 12. The digitized data from the analogue/digital converter 11 are stored in a data memory 13 that also can be wholly or partially identical to the program memory 12.

Before carrying out the voice recognition, the processor 9 corrects the sound signal picked up by the microphone 4. The correction is made on the basis of the received sound or audio information from the additional device 16. For example, the processor 9 can attempt, initially by cross-correlation, to determine the degree of correlation between the sound signal picked up by the microphone 4 and the audio signals

corresponding to the sound information from the additional device 16. In accordance with the correlation coefficients determined in this way, and the acoustic delay that can be derived between the sound signal from the microphone 4 and the audio signals corresponding to the sound information from the additional device 16, the sound signal picked up by the microphone 4 is then corrected or filtered in order to eliminate the contribution of the audio signals corresponding to the sound information. In this case, each audio signal

from the individual sound sources is subtracted from the sound  
signal from the microphone 4. As shown in Fig. 1, in the case  
of a 5-channel multi-channel sound amplifier 22 such as those  
sold under the trademark DOLBY SURROUND®, five separate audio  
5 signals are accordingly described by the sound information  
transmitted by the additional device 16 and subsequently have  
to be subtracted separately by the processor 9 from the sound  
signal from the microphone 4. Through this subtraction, only  
a sound signal in which the audio signals corresponding to the  
10 sound information have been suppressed remains. Therefore, a  
voice command contained in the filtered or corrected sound  
signal is only still distorted by an error component  
consisting of predominantly unknown interfering noise.

The processor 9 then subjects the sound signal conditioned in  
15 this way to a voice recognition algorithm (for example HMM  
(Hidden Markov Model) or DTW (Dynamic Time Warping)). These  
algorithms compare the digitized and corrected sound signal  
with predefined patterns. If the agreement in accordance with  
the algorithm respectively used is adequate to designate a  
20 pattern as identified, a process associated with the  
recognized pattern is started in the processor 9, as a result  
of which, a predefined control command is transmitted to the  
additional device 16 via a transmitter amplifier 14 and an IR  
transmitting diode 6. In this case, this may be a predefined

IR pulse train that has previously been programmed into the remote control 1.

Instead of only one microphone 4, the use of a number of microphones 4 is also conceivable. These microphones can be coupled to one another, for example, by evaluating the acoustic phase shift present between these microphones. This coupling can be carried out both before and after the correction or filtering of the sound signal. Through coupling, the physical range from which an acoustic input is intended to be permissible can be restricted accordingly.

If already existing items of equipment or systems with remote control are to be expanded by a voice recognition function, or a voice recognition function already present in the remote control 1 is to be programmed, the control commands generated by the remote control 1 should be placed into a relationship with the corresponding voice commands. To this end, the remote control 1 can be changed into a learning mode. Learning mode on the remote control 1 can be entered, for example, by pressing a button on the keyboard 2 of the remote control or a voice command (if this can already be recognized). Subsequently, a desired control function of the remote control 1 is selected, and the user is requested, for example via a small LCD display 3 on the remote control, to

input a suitable speech pattern. The input of the speech pattern can then be carried out by means of repeated recitation and recording of the voice command, by using a so-called say-in tool on the remote control 1, by inputting a phoneme sequence via the keyboard 2 or by selecting predefined words and combinations thereof to form the desired voice command, and so on. The input can be terminated by pressing the button again or by a suitable voice command.

In the example shown in Figure 1, the additional device 16 and the loudspeakers 24 are provided jointly for the television set 21 and the stereo system 20, 22. However, the individual sound sources can be operated with respectively individual loudspeakers and/or individual additional devices 16. In this case, each additional device transmits to the remote control 1 only sound information relating to the audio or sound signal produced by the corresponding sound source at that instant, and only the control commands or control information determined by the remote control 1 for the corresponding sound source being evaluated.